View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by University of Essex Research Repositor

RESEARCH ARTICLE

SOCIAL SCIENCES

2016 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

10.1126/sciadv.1600451

Humans display a reduced set of consistent behavioral phenotypes in dyadic games

**Julia Poncela-Casasnovas,[1] Mario Gutiérrez-Roig,[2] Carlos Gracia-Lázaro,[3] Julian Vicens,[1,4] Jesús Gómez-Gardeñes,[3,5] Josep Perelló,[2,6] Yamir Moreno,[3,7,8] Jordi Duch,[1] Angel Sánchez[3,9,10]*

Socially relevant situations that involve strategic interactions are widespread among animals and humans alike. To study these situations, theoretical and experimental research has adopted a game theoretical perspective, generating valuable insights about human behavior. However, most of the results reported so far have been obtained from a population perspective and considered one specific conflicting situation at a time. This makes it difficult to extract conclusions about the consistency of individuals' behavior when facing different situations and to define a comprehensive classification of the strategies underlying the observed behaviors. We present the results of a lab-in-the-field experiment in which subjects face four different dyadic games, with the aim of establishing general behavioral rules dictating individuals' actions. By analyzing our data with an unsupervised clustering algorithm, we find that all the subjects conform, with a large degree of consistency, to a limited number of behavioral phenotypes (envious, optimist, pessimist, and trustful), with only a small fraction of undefined subjects. We also discuss the possible connections to existing interpretations based on a priori theoretical approaches. Our findings provide a relevant contribution to the experimental and theoretical efforts toward the identification of basic behavioral phenotypes in a wider set of contexts without aprioristic assumptions regarding the rules or strategies behind actions. From this perspective, our work contributes to a fact-based approach to the study of human behavior in strategic situations, which could be applied to simulating societies, policy-making scenario building, and even a variety of business applications.

INTRODUCTION

Many situations in life entail social interactions where the parties involved behave strategically; that is, they take into consideration the anticipated responses of actors who might otherwise have an impact on an outcome of interest. Examples of these interactions include social dilemmas where individuals face a conflict between self and collective interests, which can also be seen as a conflict between rational and irrational decisions (*1*–*3*), as well as coordination games where all parties are rewarded for making mutually consistent decisions (*4*). These and related scenarios are commonly studied in economics, psychology, political science, and sociology, typically using a game theoretic framework to understand how decision-makers approach conflict and cooperation under highly simplified conditions (*5*–*7*).

Extensive work has shown that, when exposed to the constraints introduced in game theory designs, people are often not "rational" in the sense that they do not pursue exclusively self-interested objectives (*8*, *9*). This is especially clear in the case of prisoner's dilemma (PD) games, where rational choice theory predicts that players will always defect but empirical observation shows that cooperation oftentimes occurs, even in "one-shot" games where there is no expectation of future interaction among the parties involved (*8*, *10*). These findings beg the question as to why players sometimes choose to cooperate despite incentives not to do so. Are these choices a function of a person's identity and therefore consistent across different strategic settings? Do individuals draw from a small repertoire of responses, and if so, what are the conditions that lead them to choose one strategy over another?

Here, we attempt to shed light on these questions by focusing on a wide class of simple dyadic games that capture two important features of social interaction, namely, the temptation to free-ride and the risk associated with cooperation (*8*, *11*, *12*). All are two-person, two-action games in which participants decide simultaneously which of the two actions they will take. Following previous literature, we classify participants' set of choices as either cooperation, which we define as a choice that promotes the general interest, or defection, a choice that serves an actor's self-interest at the expense of others.

The games used in our study include PD (*13*, *14*), the stag hunt (SH) (*4*), and the hawk-dove (*15*) or snowdrift (*16*) games (SGs). SH is a coordination game in which there is a risk in choosing the best possible option for both players: cooperating when the other party defects poses serious consequences for the cooperator, whereas the defector faces less extreme costs for noncooperation (*17*). SG is an anticoordination game where one is tempted to defect, but participants face the highest penalties if both players defect (*18*). In PD games, both tensions are present: when a player defects, the counterpart faces the worst possible situation if he or she cooperates, whereas in that case, the defector benefits more than by cooperating. We also consider the harmony game (HG), where the best individual and collective options coincide; therefore, there should be no tensions present (*19*).

Several theoretical perspectives have sought to explain the seemingly irrational behavior of actors during conflict and cooperation games.

[1]Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, 43007 Tarragona, Spain. [2]Departament de Física de la Matèria Condensada, Universitat de Barcelona, 08028 Barcelona, Spain. [3]Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, 50018 Zaragoza, Spain. [4]Applied Research Group in Education and Technology, Universitat Rovira i Virgili, 43007 Tarragona, Spain. [5]Department of Condensed Matter Physics, University of Zaragoza, 50009 Zaragoza, Spain. [6]UBICS Universitat de Barcelona Institute of Complex Systems, 08028 Barcelona, Spain. [7]Department of Theoretical Physics, University of Zaragoza, 50009 Zaragoza, Spain. [8]Complex Networks and Systems Lagrange Laboratory, Institute for Scientific Interchange, 10126 Turin, Italy. [9]Grupo Interdisciplinar de Sistemas Complejos, Departamento de Matemáticas, Universidad Carlos III de Madrid, 28911 Leganés, Madrid, Spain. [10]UC3M-BS Institute of Financial Big Data, Universidad Carlos III de Madrid, 28903 Getafe, Madrid, Spain.
*Corresponding author. Email: anxo@math.uc3m.es

Perhaps most prominent among them is the theory of social value orientations (*20*–*22*), which focuses on how individuals divide resources between self and others. This research avenue has found that individuals tend to fall into certain categories such as individualistic (thinking only about themselves), competitive (attempting to maximize the difference between their own and the other's payoff), cooperative (attempting to maximize everyone's outcome), and altruistic (sacrificing their own benefits to help others). Relatedly, social preferences theory posits that people's utility functions often extend beyond their own material payoff and may include considerations of aggregate welfare or inequity aversion (*23*). Whereas theories of social orientation and social preferences assume intrinsic value differences between individuals, cognitive hierarchy theory instead assumes that players make choices on the basis of their predictions about the likely actions of other players, and as such, the true differences between individuals come not from values but rather from depth of strategic thought (*24*).

One way to arbitrate between existing theoretical paradigms is to use within-subject experiments, where participants are exposed to a wide variety of situations requiring strategic action. If individuals exhibit a similar logic (and corresponding behavior) in different experimental settings, this would provide a more robust empirical case for theories that argue that strategic action stems from intrinsic values or social orientation. By contrast, if participants' strategic behavior depends on the incentive structure afforded by the social context, these findings would pose a direct challenge to the idea that social values drive strategic choices.

We therefore contribute to the literature on decision-making in three important ways. First, we expose the same participants to multiple games with different incentive structures to assess the extent to which strategies stem from stable characteristics of an individual. Second, we depart from existing paradigms by not starting from an a priori classification to analyze our experimental data. For instance, empirical studies have typically used classifications schemes that were first derived from theory, making it difficult to determine whether these classifications are the best fit for the available data. We address this issue by using an unsupervised, robust classification algorithm to identify the full set of "strategic phenotypes" that constitute the repertoire of choices among individuals in our sample. Finally, we advance research that documents the profiles of cooperative phenotypes (*25*) by expanding the range of human behaviors that may fall into similar types of classification. By focusing on both cooperation and defection, this approach allows us to make contributions toward a taxonomy of human behaviors (*26*, *27*).

RESULTS

Laboratory-in-the-field experiment

We recruited 541 subjects of different ages, educational level, and social status during a fair in Barcelona (see Materials and Methods) (*28*). The experiment consisted of multiple rounds, in which participants were randomly assigned partners and assigned randomly chosen payoff values, allowing us to study the behavior of the same subject in a variety of dyadic games including PD, SH, SG, and HG, with different payoffs. To incentivize the experimental subjects' decisions with real material (economic) consequences, they were informed that they would proportionally receive lottery tickets (one ticket per 40 points; the modal number of tickets earned was two) to the payoff they accumulated during the rounds of dyadic games they played. The prize in the corresponding lottery was four coupons redeemable at participating neighboring stores, worth 50 euros each. The payoff matrices shown to the participants had the following form (rows are participant's strategies, whereas columns are those of the opponent)

$$\begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} R & S \\ T & P \end{pmatrix} \end{array} \quad (1)$$

Actions C and D were coded as two randomly chosen colors in the experiment to avoid framing effects. R and P were always set to $R = 10$ and $P = 5$, whereas T and S took values $T \in \{5, 6, \ldots, 15\}$ and $S \in \{0, 1, \ldots, 10\}$. In this way, the (T, S) plane can be divided into four quadrants, each one corresponding to a different game depending on the relative order of the payoffs: HG ($S > P, R > T$), SG ($T > R > S > P$), SH ($R > T > P > S$), and PD ($T > R > P > S$). Matrices were generated with equal probability for each point in the (T, S) plane, which was discretized as a lattice of 11×11 sites. Points in the boundaries between games, at the boundary of our game space, or in its center do not correspond to the four basic games previously described. However, we kept those points to add generality to our exploration, and in any event, we made sure in the analysis that the results did not change even if we removed those special games (see below). For reference, see Fig. 1 (middle) for the Nash (symmetric) equilibrium structure of each one of these games.

Population-level behavior

The average level of cooperation aggregated over all games and subjects is $\langle C \rangle = 0.49 \pm 0.01$, where the error corresponds to a 95% confidence interval (we apply this rule to the rest of our results, unless otherwise specified). This is in agreement with the theoretically expected value, $\langle C \rangle^{\text{theo}} = 0.5$, calculated by averaging over all the symmetric Nash equilibria for the (T, S) values analyzed. However, the aggregate cooperation heatmap looks very different from what would be obtained by simulating a population of players on a well-mixed scenario (compare right and central panels in Fig. 1).

On the other hand, the experimental levels of cooperation per game (excluding the boundaries between them, so the points strictly correspond to one of the four games) are as follows: $\langle C \rangle_{\text{PD}} = 0.29 \pm 0.02$ ($\langle C \rangle_{\text{PD}}^{\text{theo}} = 0$), $\langle C \rangle_{\text{SG}} = 0.40 \pm 0.02$ ($\langle C \rangle_{\text{SG}}^{\text{theo}} = 0.5$), $\langle C \rangle_{\text{SH}} = 0.46 \pm 0.02$ ($\langle C \rangle_{\text{SH}}^{\text{theo}} = 0.5$), and $\langle C \rangle_{\text{HG}} = 0.80 \pm 0.02$ ($\langle C \rangle_{\text{HG}}^{\text{theo}} = 1$). The values are considerably different from the theoretical ones in all cases, particularly for PD and HG.

Emergence of phenotypes

After looking at the behavior at the population level, we focus on the analysis of the decisions at the individual level (*27*). Our goal is to assess whether individuals behave in a highly idiosyncratic manner or whether, on the contrary, there are only a few "phenotypes" by which all our experimental subjects can be classified. To this aim, we characterize each subject with a four-dimensional vector where each dimension represents a subject's average level of cooperation in each of the four quadrants in the (T, S) plane. Then, we apply an unsupervised clustering procedure, the K-means clustering algorithm (*29*), to group those individuals that have similar behaviors, that is, the values in their vectors are similar. Input for this algorithm (see section S4.7) is the number of clusters k, which is yet to be determined, and this algorithm groups the data in such a way that it both minimizes the dispersion within

clusters and maximizes the distance among centroids of different clusters. We found that $k = 5$ clusters is the optimal number of groups according to the Davies-Bouldin index (see section S4.8) (*30*), which does not assume beforehand any specific number of types of behaviors.

The results of the clustering analysis (Fig. 2) show that there is a group that mostly cooperates in HG, a second group that cooperates in both HG and SG, and a third one that cooperates in both HG and SH. Players in the fourth group cooperate in all games, and finally, we find a small group who seems to randomly cooperate almost everywhere, with a probability of approximately 0.5.

To obtain a better understanding of the behavior of these five groups, we represent the different types of behavior in a heatmap (Fig. 3) to extract characteristic behavioral rules. In this respect, it is important to note that Fig. 3 provides a complementary view of the clustering results: our clustering analysis was carried out attending only to the aggregate cooperation level per quadrant, that is, to four numbers or coordinates per subject, whereas this plot shows the average number of times the players in each group cooperated for every point in the space of games.

The cooperation heatmaps in Fig. 3 show that there are common characteristics of subjects classified in the same group even when looking at every point of the (*T*, *S*) plane. The first two columns in Fig. 3 display consistently different behaviors in coordination and anti-coordination games, although they both act as prescribed by the Nash equlibrium in PD and HG. Both groups are amenable to a simple interpretation that links them to well-known behaviors in economic theory.

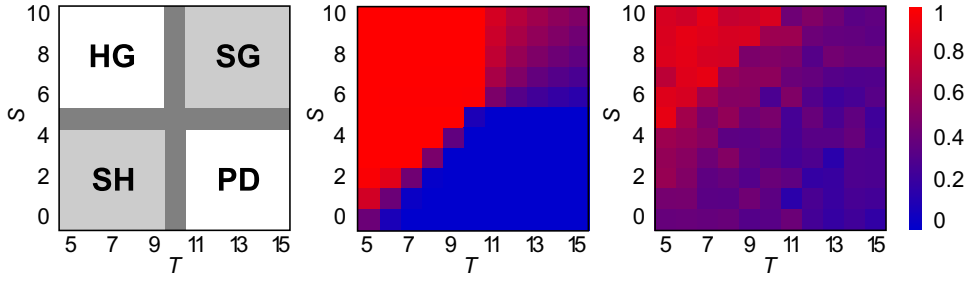

Fig. 1. Summary of the games used in the experiment and their equilibria. Schema with labels to help identify each one of the games in the quadrants of the (*T*, *S*) plane (left), along with the symmetric Nash equilibria (center) and average empirical cooperation heatmaps from the 8366 game actions of the 541 subjects (right), in each cell of the (*T*, *S*) plane. The symmetric Nash equilibria (center) for each game are as follows: PD and HG have one equilibrium, given by the pure strategies *D* and *C*, respectively. SG has a stable mixed equilibrium containing both cooperators and defectors, in a proportion that depends on the specific payoffs considered. SH is a coordination game displaying two pure-strategy stable equilibria, whose bases of attraction are separated by an unstable one, again depending on the particular payoffs of the game (*5*, *6*, *43*). The fraction of cooperation is color-coded (red, full cooperation; blue, full defection).

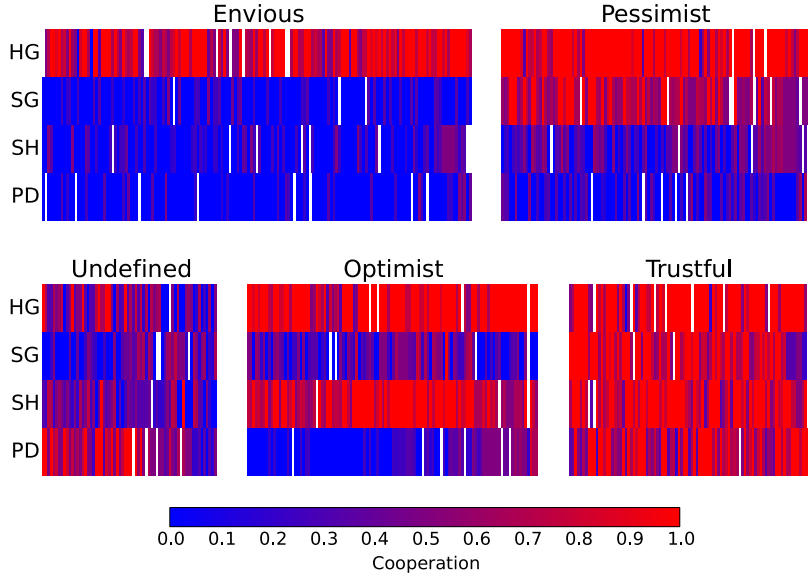

Fig. 2. Results from the *K*-means clustering algorithm. For every cluster, a column represents a player belonging to his or her corresponding cluster, whereas the four rows indicate the four average cooperation values associated with his or her (from top to bottom: cooperation in HG, SG, SH, and PD games). We color-coded the average level of cooperation for each player in each game (blue, 0.0; red, 1.0), whereas the lack of value in a particular game for a particular player is coded in white. Cluster sizes: Envious, $n = 161$ (30%); Pessimist, $n = 113$ (21%); Undefined, $n = 66$ (12%); Optimist, $n = 110$ (20%); Trustful, $n = 90$ (17%).

Thus, the first phenotype (n = 110 or 20% of the population) cooperates wherever $T < R$ (that is, they cooperate in the HG and in the SH and defect otherwise). By using this strategy, these subjects aim to obtain the maximum payoff without taking into account the likelihood that their counterpart will allow them to get it, in agreement with a maximax behavior (*31*). Accordingly, we call this first phenotype "optimists." Conversely, we label subjects in the second phenotype "pessimists" (n = 113 or 21% of the population) because they use a maximin principle (*32*) to choose their actions, cooperating only when $S > P$ (that is, in HG and SG) to ensure a best worst-case scenario. The behaviors of these two phenotypes, which can hardly be considered rational [as discussed by Colman (*31*)], are also associated with different degrees of risk aversion, a question that will be addressed below.

Regarding the third column in Fig. 3, it is apparent from the plots that individuals in this phenotype (n = 161 or 30% of the population) exclusively cooperate in the upper triangle of HG [that is, wherever $(S - T) \geq 0$]. As was the case with optimists and pessimists, this third behavior is far from being rational in a self-centered sense, in so far as players forsake the possibility of achieving the maximum payoff by playing the only Nash equilibrium in HG. In turn, these subjects seem to behave as driven by envy, status-seeking consideration, or lack of trust. By choosing D when $S > P$ and $R > T$, these players prevent their counterparts from receiving more payoff than themselves even when, by doing so, they diminish their own potential payoff. The fact that competitiveness overcomes rationality as players basically attempt to ensure they receive more payoff than their opponents suggests an interpretation of the game as an assurance game (*3*), and accordingly, we have dubbed this phenotype "envious."

The fourth phenotype (fourth column in Fig. 3) includes those players who cooperate in almost every round and in almost every site of the (T, S) plane (n = 90 or 17% of the population). In this case, and opposite to the previous one, we believe that these players' behavior can be associated with trust in partners behaving in a cooperative manner. Another way of looking at trust in this context is in terms of expectations, because it has been shown that expectation of cooperation enhances cooperation in the PD (*33*). In any event, explaining the roots of this type of cooperative behavior in a unique manner seems to be a difficult task, and alternative explanations of cooperation on the PD involving normalized measures of greed and fear (*34*) or up to five simultaneous factors (*35*) have been advanced too. Lacking an unambiguous motivation of the observed actions of the subjects in this group, we find the name "trustful" to be an appropriate one to refer to this phenotype. Last, the unsupervised algorithm found a small fifth group of players (n = 66 or 12% of the population) who cooperate in an approximately random manner, with a probability of 0.5, in any situation. For lack of better insight into their behavior, we will hereinafter refer to this minority as "undefined."

Remarkably, three of the phenotypes reported here (optimist, pessimist, and trustful) are of a very similar size. On the other hand, the largest one is the envious phenotype, including almost a third of the participants, whereas the undefined group, which we cannot yet consider as a bona fide phenotype because we have not found any interpretation of the corresponding subjects' actions, is considerably smaller than all the others. In agreement with abundant experimental evidence, we have not found any purely rational phenotype: the strategies used by the four relevant groups are, to different extents, quite far from self-centered rationality. Note that ours is an across-game characterization, which does not exclude the possibility of subjects taking rational, purely self-regarding decisions when restricted to one specific game (see section S4.5).

Finally, and to shed more light on the phenotypes found above, we estimate an indirect measure of their risk aversion. To do this, we

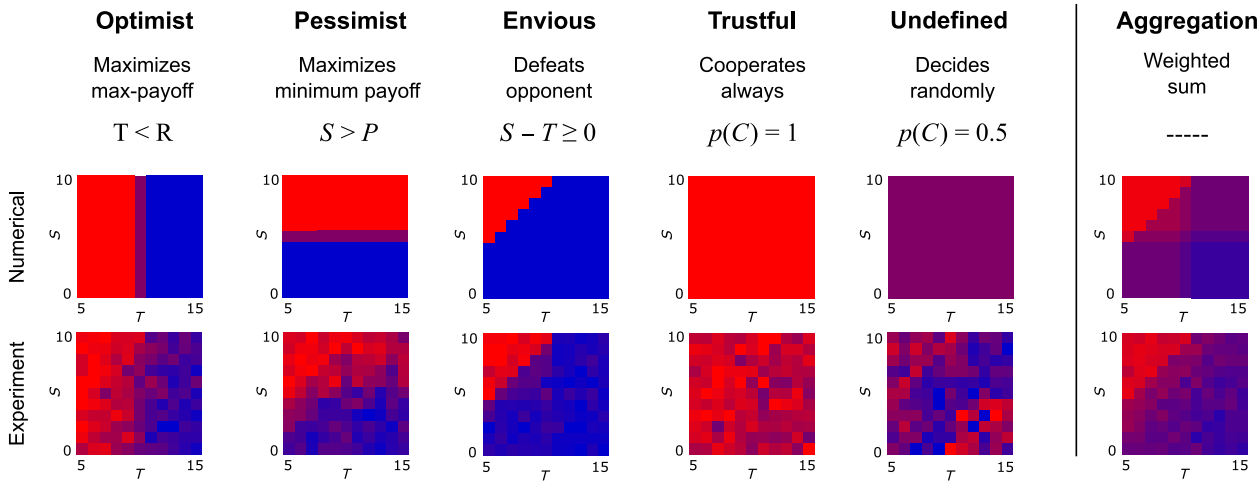

Fig. 3. Summary results of the different phenotypes (Optimist, Pessimist, Envious, Trustful, and Undefined) determined by the *K*-means clustering algorithm, plus the aggregation of all phenotypes. For each phenotype (column), we show the word description of the behavioral rule and the corresponding inferred behavior in the whole (T, S) plane (labeled as Numerical). The fraction of cooperation is color-coded (red, full cooperation; blue, full defection). The last row (labeled as Experiment) shows the average cooperation, aggregating all the decisions taken by the subjects classified in each cluster. The fractions for each phenotype are as follows: 20% Optimist, 21% Pessimist, 30% Envious, 17% Trustful, and 12% Undefined. The very last column shows the aggregated heatmaps of cooperation for both the simulations and the experimental results. The simulation results assume that each individual plays using one and only one of the behavioral rules and respects the relative fractions of each phenotype in the population found by the algorithm. Note the agreement between aggregated experimental and aggregated numerical heatmaps (the discrepancy heatmap between them is shown in section S4.11). We report that the average difference across the entire (T, S) plane between the experiment and the phenotype aggregation is of 1.39 SD units, which represents a value inside the standard 95% confidence interval, whereas for any given phenotype, this difference averaged over the entire (T, S) plane is smaller than 2.14 SD units.

consider the number of cooperative actions in SG together with the number of defective actions in SH (over the total sum of actions in both quadrants for a given player; see section S4.5). Whereas envious, trustful, and undefined players exhibit intermediate levels of risk aversion (0.52, 0.52, and 0.54, respectively), pessimists exhibit significantly higher value (0.73), consistent with their fear of facing the worst possible outcome and their choice of the best worst-case scenario. In contrast, the optimist phenotype shows a very low risk aversion (0.32), in agreement with the fact that they aim to obtain the maximum possible payoff, taking the risk that their counterpart does not work with them toward that goal.

Robustness of phenotypes

We have carefully checked that our *K*-means clustering results are robust. Lacking the "ground truth" behind our data in terms of different types of individual behaviors, we must test the significance and robustness of our clustering analysis by checking its dependence on the data set itself. We studied this issue in several complementary manners. First, we applied the same algorithm to a randomized version of our data set (preserving the total number of cooperative actions in the population but destroying any correlation among the actions of any given subject), showing no significant clustering structure at all (see section S4.7 for details).

Second, we ran the *K*-means clustering algorithm on portions of the original data with the so-called "leave-*p*-out" procedure (*36*). This test showed that the optimum five-cluster scheme found is robust even when randomly excluding up to 55% of the players and their actions (see section S4.7 for details). Moreover, we repeated the whole analysis, discarding the first two choices made by every player, to account for excessive noise due to initial lack of experience; the results more clearly show even the same optimum at five phenotypes. See section S4.7 for a complete discussion.

Third, we tested the consistency among cluster structures found in different runs of the same algorithm for a fixed number of clusters, that is to say, how likely it is that the particular composition of individuals in the cluster scheme from one realization of the algorithm is correlated with the composition from that of a different realization. To ascertain this, we computed the normalized mutual information score MI (see section S4.9 for formal definition) (*37*), knowing that the comparison of two runs with exactly the same clustering composition would give a value MI = 1 (perfect correlation), and MI = 0 would correspond to a total lack of correlation between them. We ran our *K*-means clustering algorithm 2000 times for the optimum $k = 5$ clusters and paired the clustering schemes for comparison, obtaining an average normalized mutual information score of MI = 0.97 (SD, 0.03). To put these numbers in perspective, the same score for the pairwise comparison of results from 2000 realizations of the algorithm on the randomized version of the data is MI = 0.59 (SD, 0.18) (see section S4.9 for more details).

All the tests presented above provide strong support for our classification in terms of phenotypes. However, we also searched for possible dependencies of the phenotype classification on the age and gender distributions for each group (see section S4.10), and we found no significant differences among them, which hints toward a classification of behaviors (phenotypes) beyond demographic explanations.

DISCUSSION AND CONCLUSIONS

We have presented the results of a laboratory-in-the-field experiment designed to identify phenotypes, following the terminology fittingly introduced by Peysakhovich *et al.* (*25*). Our results suggest that the individual behaviors of the subjects in our population can be described by a small set of phenotypes: envious, optimist, pessimist, trustful, and a small group of individuals referred to as undefined, who play an unknown strategy. The relevance of this repertoire of phenotypes arises from the fact that it has been obtained from experiments in which subjects played a wide variety of dyadic games through an unsupervised procedure, the *K*-means clustering algorithm, and that it is a very robust classification. With this technique, we can go beyond correlations and assign specific individuals to specific phenotypes, instead of looking at (aggregate) population data. In this respect, the trimodal distributions of the joint cooperation probability found by Capraro *et al.* (*38*) show much resemblance to our findings, and although a direct comparison is not possible because they correspond to aggregate data, they point in the direction of a similar phenotype classification. In addition, our results contribute to the currently available evidence that people are heterogeneous, by quantifying the degree of heterogeneity, in terms of both the number of types and their relative frequency, in a specific (but broad) suite of games.

Although the robustness of our agnostic identification of phenotypes makes us confident of the relevance of the behavioral classification, and our interpretation of it is clear and plausible, it is not the only possible one. It is important to point out that connections can also be drawn to earlier attempts to classify individual behaviors. As we have mentioned previously, one theory that may also shed light on our classification is that of social value orientation (*20*–*22*). Thus, the envious type may be related to the competitive behavior found in that context (although in our observation, envious people just aim at making more profit than their competitors, not necessarily minimizing their competitors' profit); optimists could be cooperative, and trustful seem very close to altruistic. As for the pessimist phenotype, we have not been able to draw a clear relationship to the types most commonly found among social value orientations, but in any event, the similarity between the two classifications is appealing and suggests an interesting line for further research. Another alternative view on our findings arises from social preferences theory (*23*), where, for instance, envy can be understood as the case in which inequality that is advantageous to self yields a positive contribution to one's utility (*39*–*42*). Altruists can be viewed as subjects with concerns for social welfare (*39*), whereas other phenotypes are difficult to be understood in this framework, and optimists and pessimists do not seem to care about their partner's outcome. However, other interpretations may apply to these cases: optimists could be players strongly influenced by payoff dominance a la Harsanyi and Selten (*43*), in the sense that these players would choose strategies associated with the best possible payoff for both. Yet, another view on this phenotype is that of team reasoning (*44*–*46*), namely, individuals whose strategies maximize the collective payoff of the player pair if this strategy profile is unique. Proposals such as the cognitive hierarchy theory (*24*, *47*) and the level-k theory (*48*, *49*) do not seem to fit our results in so far as the best response to the undefined phenotype, which would be the zeroth level of behavior, does not match any of our behavioral classes.

Our results open the door to making relevant advances in a number of directions. For instance, they point to the independence of the phenotypic classification of age and gender. Although the lack of gender dependence may not be surprising, it would be really astonishing that small children would exhibit behaviors with similar classifications in view of the body of experimental evidence about their differences from adults (*50*–*55*), and further research is needed to assess this issue in

detail. As discussed also by Peysakhovich *et al.* (*25*), our research does not illuminate whether the different phenotypes are born, made, or something in between, and thus, understanding their origin would be a far-reaching result.

We believe that applying an approach similar to ours to obtain results about the cooperative phenotype (*25*, *38*, *56*) and, even better, to carry out experiments with an ample suite of games, as well as a detailed questionnaire (*57*), is key in future research. In this regard, it has to be noted that the relationship between our automatically identified phenotypes and theories of economic behavior yields predictions about other games: envy and expectations about the future and about other players will dictate certain behaviors in many other situations. Therefore, our classification here can be tested and refined by looking for phenotypes arising in different contexts. This could be complemented with a comparison of our unsupervised algorithm with the parametric modeling approach by Cabrales (*41*) or even by implementing flexible specifications to social preferences (*23*, *39*, *40*) or social value orientation (*20*–*22*) to improve the understanding of our behavioral phenotypes.

Finally, our results also have implications in policy-making and real-life economic interactions. For instance, there is a large group of individuals, the envious ones (about a third of the population), that in situations such as HG fail to cooperate when they are at risk of being left with lower payoff than their counterpart. This points to the difficulty of making people understand when they face a nondilemmatic, win-win situation, and that effort must be expended to make this very clear. Other interesting subpopulations are those of the pessimist and optimist phenotypes, which together amount to approximately half of the population. These people exhibit large or small risk aversion, respectively, and use an ego-centered approach in their daily lives, thus ignoring that others can improve or harm their expected benefit with highly undesirable consequences. A final example of the hints provided by our results is the existence of an unpredictable fraction of the population (undefined) that, even being small, can have a strong influence on social interactions because its noisy behavior could lead people with more clear heuristics to mimic its erratic actions. On the other hand, the classification in terms of phenotypes (particularly if, as we show here, it comprises only a few different types) can be very useful for firms, companies, or banks interacting with people: it could be used to evaluate customers or potential ones or even employees for managerial purposes, allowing for a more efficient handling of the human resources in large organizations. This approach is also very valuable in the emergent deliberative democracy and open-government practices around the globe [including the Behavioural Insights Team (*58*) of the UK government, its recently established counterpart at the White House or the World Health Organization (*59*)]. Research following the lines presented here could lead to many innovations in these contexts.

MATERIALS AND METHODS

The experiment was conducted as a lab-in-the-field, that is, to avoid restricting ourselves to the typical samples of university undergraduate students, we took our laboratory to a festival in Barcelona and recruited subjects from the general audience (*28*). This setup allows, at the very least, to obtain results from a very wide age range, as was the case in a previous study where it was found that teenagers behave differently (*55*). All participants in the experiment signed an informed consent to participate. In agreement with the Spanish Law for Personal Data Protection, no association was ever made between their real names and the results. This procedure was checked and approved by the Vice-provost of Research of Universidad Carlos III de Madrid, the institution funding the experiment.

To equally cover the four dyadic games in our experiments, we discretized the (T, S) plane as a lattice of 11×11 sites. Each player was equipped with a tablet running the application of the experiment (see section S1 for technical details and section S2 for the experiment protocol). The participants were shown a brief tutorial in the tablet (see the translation of the tutorial in section S3) but were not instructed in any particular way nor with any particular goal in mind. They were informed that they had to make decisions in different conditions and against different opponents in every round. They were not informed about how many rounds of the game they were going to play. Because of practical limitations, we could only simultaneously host around 25 players, so the experiment was conducted in several sessions over a period of 2 days. In every session, all individuals played a different, randomly picked number of rounds between 13 and 18. In each round of a session, each participant was randomly assigned a different opponent and a payoff matrix corresponding to a different (T, S) point among our 11×11 different games. Couples and payoff matrices were randomized in each new round, and players did not know the identity of their opponents. In case there was an odd number of players or a given player was nonresponsive, the experimental software took over and made the game decision for him or her, accordingly labeling its corresponding data to discard actions in the analysis (143 actions). When the action was actually carried out by the software, the stipulation was that it repeated the previous choice of C or D with an 80% probability. In the three cases where a session had an odd number of participants, it has to be noted that no subjects played all the time against the software, because assigning of partners was randomized for every round. The total number of participants in our experiment was 541, adding up to a total of 8366 game decisions collected, with an average number of actions per (T, S) value of 69.1 (see also section S4.3).

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/2/8/e1600451/DC1

Technical implementation of the experiment

Running the experiment

Translated transcript of the tutorial and feedback screen after each round

Other experimental results

fig. S1. System architecture.

fig. S2. Age distribution of the participants in our experiment.

fig. S3. Screenshots of the tutorial shown to participants before starting the experiment and feedback screen after a typical round of the game.

fig. S4. Fraction of cooperative actions for young (≤15 years old) and adult players (>16 years old) and relative difference between the two heatmaps: (young − adults)/adults.

fig. S5. Fraction of separate cooperative actions for males and females and relative difference between the two heatmaps: (males − females)/females.

fig. S6. Fraction of cooperative actions separated by round number: for the first 1 to 3 rounds, 4 to 10 rounds, and last 11 to 18 rounds.

fig. S7. Relative difference in the fraction of cooperation heatmaps between groups of rounds.

fig. S8. Total number of actions in each point of the (T,S) plane for all 541 participants in the experiment (the total number of game actions in the experiment adds up to 8366).

fig. S9. SEM fraction of cooperative actions in each point of the (T,S) plane for all the participants in the experiment.

fig. S10. Average fraction of cooperative actions (and SEM) among the population as a function of the round number overall (left) and separating the actions by game (right).

fig. S11. Distribution of fraction of rational actions among the 541 subjects of our experiment, when considering only their actions in HG or PD, or both.

fig. S12. Fraction of rational actions as a function of the round number for the 541 subjects, defined by their actions in the PD game and HG together (top) and independently (bottom).
fig. S13. Values of risk aversion averaged over the subjects in each phenotype.
fig. S14. Average response times (and SEM) as a function of the round number for all the participants in the experiment and separating the actions into cooperation or defection.
fig. S15. Distributions of response times for all the participants in the experiment and separating the actions into cooperation (top) and defection (bottom).
fig. S16. Testing the robustness of the results from the *K*-means algorithm.
fig. S17. Davies-Bouldin index as a function of the number of clusters in the partition of our data (dashed black) compared to the equivalent results for different leave-*p*-out analyses.
fig. S18. Average value for the normalized mutual information score, when doing pairwise comparisons of the clustering schemes from 2000 independent runs of the *K*-means algorithm both on the actual data and on the randomized version of the data.
fig. S19. Age distribution for the different phenotypes compared to the distribution of the whole population (black).
fig. S20. Difference between the experimental (second row) and numerical (or inferred; first row) behavioral heatmaps for each one of the phenotypes found by the *K*-means clustering algorithm, in units of SD.
fig. S21. Average level of cooperation over all game actions and for different values of *T* (in different colors).
fig. S22. Average level of cooperation as a function of (*T*,*S*) for both hypothesis and experiment.

REFERENCES AND NOTES

1. R. M. Dawes, Social dilemmas. *Annu. Rev. Psychol.* **31**, 169–193 (1980).
2. P. Kollock, Social dilemmas: The anatomy of cooperation. *Annu. Rev. Soc.* **24**, 183–214 (1998).
3. P. A. M. Van Lange, J. Joireman, C. D. Parks, E. Van Dijk, The psychology of social dilemmas: A review. *Organ. Behav. Hum. Dec. Process.* **120**, 125–141 (2013).
4. B. Skyrms, *The Stag Hunt and the Evolution of Social Structure* (Cambridge Univ. Press, Cambridge, UK, 2003).
5. K. Sigmund, *The Calculus of Selfishness* (Princeton Univ. Press, Princeton, NJ, 2010).
6. H. Gintis, *Game Theory Evolving: A Problem-centered Introduction to Evolutionary Game Theory* (Princeton Univ. Press, Princeton, NJ, ed. 2, 2009).
7. R. B. Myerson, *Game Theory—Analysis of Conflict* (Harvard Univ. Press, Cambridge, MA, 1991).
8. C. F. Camerer, *Behavioral Game Theory: Experiments in Strategic Interaction* (Princeton Univ. Press, Princeton, NJ, 2003).
9. J. H. Kagel, A. E. Roth, *The Handbook of Experimental Economics* (Princeton Univ. Press, Princeton, NJ, 1997).
10. J. O. Ledyard, Public goods: A survey of experimental research, in *The Handbook of Experimental Economics*, J. H. Kagel, A. E. Roth, Eds. (Princeton Univ. Press, Princeton, NJ, 1997), pp. 111–194.
11. A. Rapoport, M. Guyer, A taxonomy of 2 × 2 games. *Gen. Syst.* **11**, 203–214 (1966).
12. M. W. Macy, A. Flache, Learning dynamics in social dilemmas. *Proc. Natl. Acad. Sci. U.S.A.* **99** (suppl. 3), 7229–7236 (2002).
13. A. Rapoport, A. M. Chammah, *Prisoner's Dilemma* (University of Michigan Press, Ann Arbor, MI, 1965).
14. R. Axelrod, W. D. Hamilton, The evolution of cooperation. *Science* **211**, 1390–1396 (1981).
15. J. M. Smith, *Evolution and the theory of games* (Cambridge Univ. Press, Cambridge, UK, 1982).
16. R. Sugden, *The Economics of Rights, Cooperation and Welfare* (Palgrave Macmillan, London, UK, ed. 2, 2005).
17. R. Cooper, *Coordination Games* (Cambridge Univ. Press, Cambridge, UK, 1998).
18. Y. Bramoullé, Anti-coordination and social interactions. *Games Econ. Behav.* **58**, 30–49 (2007).
19. A. N. Licht, Games commissions play: 2×2 Games of international securities regulation. *Yale J. Int. Law* **24**, 61–125 (1999).
20. P. M. A. Van Lange, Beyond self-interest: A set of propositions relevant to interpersonal orientations. *Eur. Rev. Soc. Psychol.* **11**, 297–331 (2000).
21. C. E. Rusbult, P. A. M. Van Lange, Interdependence, interaction, and relationships. *Annu. Rev. Psychol.* **54**, 351–375 (2003).
22. D. Balliet, C. Parks, J. Joireman, Social value orientation and cooperation in social dilemmas: A meta-analysis. *Group Process. Interg. Rel.* **12**, 533–547 (2009).
23. E. Fehr, K. M. Schmidt, A theory of fairness, competition, and cooperation. *Q. J. Econ.* **114**, 817–868 (1999).
24. C. F. Camerer, T.-H. Ho, J.-K. Chong, A cognitive hierarchy model of games. *Q. J. Econ.* **119**, 861–898 (2004).
25. A. Peysakhovich, M. A. Nowak, D. G. Rand, Humans display a 'cooperative phenotype' that is domain general and temporally stable. *Nat. Commun.* **5**, 4939 (2014).
26. M. Blanco, D. Engelmann, H. T. Normann, A within-subject analysis of other-regarding preferences. *Games Econ. Behav.* **72**, 321–338 (2011).
27. A. P. Kirman, Whom or what does the representative individual represent? *J. Econ. Perspec.* **6**, 117–136 (1992).
28. O. Sagarra, M. Gutiérrez-Roig, I. Bonhoure, J. Perelló, Citizen science practices for computational social science research: The conceptualization of pop-up experiments. *Front. Phys.* **3**, 93 (2016).
29. J. MacQueen, Some methods for classification and analysis of multivariate observations, in *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability* (University of California Press, Berkeley, CA, 1967), pp. 281–297.
30. D. L. Davies, D. W. Bouldin, A cluster separation measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **1**, 224–227 (1979).
31. A. M. Colman, *Game Theory and its Applications: In the Social and Biological Sciences* (Psychology Press, Routledge, Oxford, UK, 1995).
32. J. Von Neumann, O. Morgenstern, *Theory of Games and Economic Behavior* (Princeton Univ. Press, Princeton, NJ, 1944).
33. G. T. T. Ng, W. T. Au, Expectation and cooperation in prisoner's dilemmas: The moderating role of game riskiness. *Psychon. Bull. Rev.* **23**, 353–360 (2016).
34. T. K. Ahn, E. Ostrom, D. Schmidt, R. Shupp, J. Walker, Cooperation in PD games: Fear, greed, and history of play. *Public Choice* **106**, 137–155 (2001).
35. C. Engel, L. Zhurakhovska, "When is the risk of cooperation worth taking? The prisoner's dilemma as a game of multiple motives" (Max Planck Institute for Research on Collective Goods no. 2012/16, Bonn, 2012).
36. R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection. *IJCAI* **14**, 1137–1145 (1995).
37. D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms* (Cambridge Univ. Press, Cambridge, UK, ed. 2, 2003).
38. V. Capraro, J. J. Jordan, D. G. Rand, Heuristics guide the implementation of social preferences in one-shot Prisoner's Dilemma experiments. *Sci. Rep.* **4**, 6790 (2014).
39. G. Charness, M. Rabin, Understanding social preferences with simple tests. *Q. J. Econ.* **117**, 817–869 (2002).
40. G. Bolton, A. Ockenfels, ERC: A theory of equity, reciprocity and competition. *Am. Econ. Rev.* **90**, 166–193 (2000).
41. A. Cabrales, The causes and economic consequences of envy. *SERIEs* **1**, 371–386 (2010).
42. A. Cabrales, R. Miniaci, M. Piovesan, G. Ponti, Social preferences and strategic uncertainty: An experiment on markets and contracts. *Am. Econ. Rev.* **100**, 2261–2278 (2010).
43. J. C. Harsanyi, R. Selten, *A General Theory of Equilibrium Selection in Games* (Massachusetts Institute of Technology Press, Cambridge, MA, 1988).
44. M. Bacharach, Interactive team reasoning: A contribution to the theory of co-operation. *Res. Econ.* **53**, 117–147 (1999).
45. R. Sugden, Thinking as a team: Towards an explanation of nonselfish behaviour. *Soc. Philos. Policy* **10**, 69–89 (1993).
46. R. Sugden, Mutual advantage, conventions and team reasoning. *Int. Rev. Econ.* **58**, 9–20 (2011).
47. A. M. Colman, B. D. Pulford, C. L. Lawrence, Explaining strategic coordination: Cognitive hierarchy theory, strong Stackelberg reasoning, and team reasoning. *Decision* **1**, 35–58 (2014).
48. D. O. Stahl II, P. W. Wilson, Experimental evidence on players' models of other players. *J. Econ. Behav. Organ.* **25**, 309–327 (1994).
49. D. O. Stahl, P. W. Wilson, On players' models of other players: Theory and experimental evidence. *Games Econ. Behav.* **10**, 218–254 (1995).
50. E. Fehr, H. Bernhard, B. Rockenbach, Egalitarianism in young children. *Nature* **54**, 1079–1083 (2008).
51. B. House, J. Henrich, B. Sarnecka, J. B. Silk, The development of contingent reciprocity in children. *Evol. Hum. Behav.* **34**, 86–93 (2013).
52. G. Charness, M.-C. Villeval, Cooperation and competition in intergenerational experiments in the field and the laboratory. *Am. Econ. Rev.* **99**, 956–978 (2009).
53. M. Sutter, M. G. Kocher, Trust and trustworthiness across different age groups. *Games Econ. Behav.* **59**, 364–382 (2007).
54. J. F. Benenson, J. Pascoe, N. Radmore, Children's altruistic behavior in the dictator game. *Evol. Hum. Behav.* **28**, 168–175 (2007).
55. M. Gutiérrez-Roig, C. Gracia-Lázaro, J. Perelló, Y. Moreno, A. Sánchez, Transition from reciprocal cooperation to persistent behaviour in social dilemmas at the end of adolescence. *Nat. Commun.* **5**, 4362 (2014).
56. T. Yamagishi, N. Mifune, Y. Li, M. Shinada, H. Hashimoto, Y. Horita, A. Miura, K. Inukai, S. Tanida, T. Kiyonari, H. Takagishi, D. Simunovic, Is behavioral pro-sociality game-specific? Pro-social preference and expectations of pro-sociality. *Org. Behav. Human Decis. Proc.* **120**, 260–271 (2013).
57. F. Exadaktylos, A. M. Espín, P. Brañas-Garza, Experimental subjects are not different. *Sci. Rep.* **3**, 1213 (2013).
58. The Behavioural Insights Team, www.behavioural insights.co.uk.
59. World Health Organization, www.who.int/topics/obesity/en.

Acknowledgments: We thank P. Brañas-Garza, A. Cabrales, A. Espín, A. Hockenberry, and A. Pah, as well as our two anonymous reviewers, for their useful comments. We thank K. Gaughan for his thorough grammar and editing suggestions. We also acknowledge the participation of 541 anonymous volunteers who made this research possible. We are indebted to the BarcelonaLab program through the Citizen Science Office promoted by the Direction of Creativity and Innovation of the Institute of Culture of the Barcelona City Council led by I. Garriga for their help and support for setting up the experiment at the Dau Barcelona Festival at Fabra i Coats. We specially want to thank I. Bonhoure, O. Marín from Outliers, N. Fernández, C. Segura, C. Payrató, and P. Lorente for all the logistics in making the experiment possible and to O. Comas (director of the DAU) for giving us this opportunity. **Funding:** This work was partially supported by Mineco (Spain) through grants FIS2013-47532-C3-1-P (to J.D.), FIS2013-47532-C3-2-P (to J.P.), FIS2012-38266-C02-01 (to J.G.-G.), and FIS2011-25167 (to J.G.-G. and Y.M.); by Comunidad de Aragón (Spain) through the Excellence Group of Non Linear and Statistical Physics (FENOL) (to C.G.-L., J.G.-G., and Y.M.); by Generalitat de Catalunya (Spain) through Complexity Lab Barcelona (contract no. 2014 SGR 608; to J.P. and M.G.-R.) and through Secretaria d'Universitats i Recerca (contract no. 2013 DI 49; to J.D. and J.V.); and by the European Union through Future and Emerging Technologies FET Proactive Project MULTIPLEX (Multilevel Complex Networks and Systems) (contract no. 317532; to Y.M., J.G.-G., and J.P.-C.) and FET Proactive Project DOLFINS (Distributed Global Financial Systems for Society) (contract no. 640772; to C.G.-L., Y.M., and A.S.). **Author contributions:** J.P., Y.M., and A.S. conceived the original idea for the experiment; J.P.-C., C.G.-L., J.V., J.G.-G., J.P., Y.M., J.D., and AS contributed to the final experimental setup; J.V., J.D., and J.P.-C. wrote the software interface for the experiment; J.P.-C., M.G.-R., C.G.-L., J.G.-G., J.P., Y.M., and J.D. carried out the experiments; J.P.-C., M.G.-R., C.G.-L., and J.G.-G. analyzed the data; J.P.C., M.G.-R., C.G.-L., J.G.-G., J.P., Y.M., J.D., and A.S. discussed the analysis results; and J.P.-C., M.G.-R., C.G.-L., J.V., J.G.-G., J.P., Y.M., J.D., and A.S. wrote the paper. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 1 March 2016
Accepted 2 July 2016
Published 5 August 2016
10.1126/sciadv.1600451

Citation: J. Poncela-Casasnovas, M. Gutiérrez-Roig, C. Gracia-Lázaro, J. Vicens, J. Gómez-Gardeñes, J. Perelló, Y. Moreno, J. Duch, A. Sánchez, Humans display a reduced set of consistent behavioral phenotypes in dyadic games. *Sci. Adv.* **2**, e1600451 (2016).